

AD-A152 214

ON THE CHOICE OF COORDINATES FOR SEMICONDUCTOR
SIMULATION(U) YALE UNIV NEW HAVEN CT DEPT OF COMPUTER
SCIENCE T KERKHOVEN DEC 84 YALEU/DCS/RR-350

1/1

UNCLASSIFIED

N00014-82-K-0184

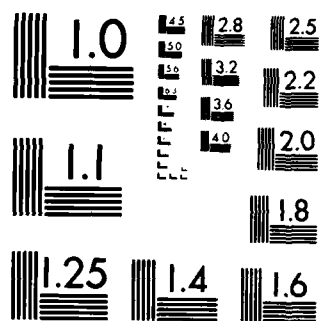
F/G 20/12

NL

END

FILMED

PTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

AD-A152 214



On the Choice of Coordinates for Semiconductor Simulation

Tom Kerkhoven

Research Report YALEU/DCS/RR-350

December 1984

This document has been approved
for public release and sale; its
distribution is unlimited.

APR 04 1985

DTIC FILE COPY

YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

85 03 13 242

REPRODUCED AT GOVERNMENT EXPENSE

On the Choice of Coordinates for Semiconductor Simulation

Tom Kerkhoven

**Research Report YALEU/DCS/RR-350
December 1984**

The work presented in this paper was supported in part by the Department of Energy under contract DE-ACO2-81ER10996, in part by the Office of Naval Research under grant N00014-82-K-0184 and in part by the US Army Research Office under grant DAAG-83-0177.

Abstract

The semiconductor simulation problem is defined by a set of three coupled nonlinear elliptic partial differential equations. These equations are usually solved either by a coupled or by a decoupling algorithm. Because the solution of semiconductor simulation problems varies over many orders of magnitude, the numerical solution of this nonlinear problem easily gives rise to scaling and conditioning problems in either case.

In this report it is shown that important properties of the numerical solution depend on the method of discretization and on three systems of coordinates: The system of coordinates which is used for discretization, the system of coordinates in which we linearize and possibly a third system of coordinates employed for the solution of the linear systems of equations. For the solution to the original partial differential equations charge is conserved and certain maximum principles are valid. In order to preserve these properties for the discretized solution special care has to be taken with respect to the discretization procedure and the choice of coordinates. It is shown that the choice of solution algorithm and the desire to preserve charge and to obtain acceptable numerical properties restrict the possible discretization procedures and sets of coordinates.

Accession No. *X*

LCIS 001 I *per*

1000 1000

ltr

A-1



1. Introduction

In this report we consider the usual macroscopic analysis of the flow of electrons and holes in a semiconductor which is described in terms of the original physical variables u, n and p (see below) by the system of nonlinear partial differential equations (pdes) as originally given by van Roosbroeck in [16]. The physical origin of these equations is discussed concisely in [3]. A more extensive treatment of the relevant physics can be found in [18].

Because the mathematical equations modeling the charge transport in a semiconductor device are relatively complicated, extensive simplifying assumptions have to be made in order to make analytical treatment at all feasible. The analytical models are therefore not sufficiently accurate for the simulation of modern micro-technological semiconductor devices and numerical solution methods have to be employed.

For the solution to the original system of pdes, physical charge is conserved and certain maximum principles are valid. We show how the desire to retain the corresponding properties for the discretized equations determines a number of aspects of the discretization procedure and the system of coordinates for the case of a tensor product mesh. The approach generalizes straightforwardly to more general meshes, however.

We advocate a discretization procedure which results in discretized nonlinear equations which satisfy the properties mentioned above and show that numerical solution of these equations is problematic. In fact, it turns out that it is imperative to transform the discretized equations with the desired theoretical properties to a second system of coordinates with more favorable properties for numerical solution. This can be done without sacrificing the acquired theoretical properties. The set of nonlinear discretized equations which is thus obtained is solved iteratively by repeatedly solving linearized systems of equations as usual. We may want to improve the conditioning of these linear systems of equations by the introduction of a third set of coordinates. Thus we develop a more systematic treatment of the discretization process and the choice of coordinates than has been given in earlier publications such as [1].

In § 2 we summarize the physical equations and the boundary conditions for the problem. In § 3 we examine the original system of partial differential equations and show that charge is conserved while certain maximum principles hold for the solution. In § 4 we introduce a discretization procedure and show that charge can be conserved for the discrete problem while the appropriate discrete maximum principles can be satisfied. In § 5 the importance of the choice of coordinates for solution of the equations is discussed. In § 6 we argue that for numerical solution we want a

representation of the simulation problem which differs from the one used in section 4. In § 7 we show that we can obtain suitable equations in terms of the original physical coordinates: Potential and densities. In § 8 we derive equations in terms of the so-called quasi Fermi levels, which are smoother than the original coordinates. In § 9 we discuss alternative formulations and show that these are hardly useful.

2. Mathematical description of semiconductors

The Equations. The device geometry for a typical two dimensional model of an n-channel MOSFET is shown in figure 1. The actual semiconductor region of the device in which the charge transport occurs, is given by the major quadrangle Ω with boundary A-B-C-F-G-H. Electric potentials are applied at the source, gate, drain and backgate contacts. As is indicated in the figure, the so-called gate contact is separated from the semiconductor medium by a thin oxide layer. The region with boundary A-B-C-D-E-F-G-H which includes the thin oxide layer on top, is called Ω' .

The first partial differential equation describing the intrinsic behaviour of a semiconductor device is Maxwell's equation for the electrostatic potential ψ as a function of the charge density ρ :

$$-\nabla \cdot (\epsilon_s \epsilon_0 \nabla \psi) = \rho.$$

Within the semiconductor region Ω the total space-charge ρ is given by

$$\rho = -q(n - p - k_1),$$

where

q is the size of the charge of the electron,

n is the electron charge density,

p is the hole charge density,

ϵ_0 is the dielectric constant of the vacuum,

ϵ_s is the relative dielectric constant of the semiconductor,

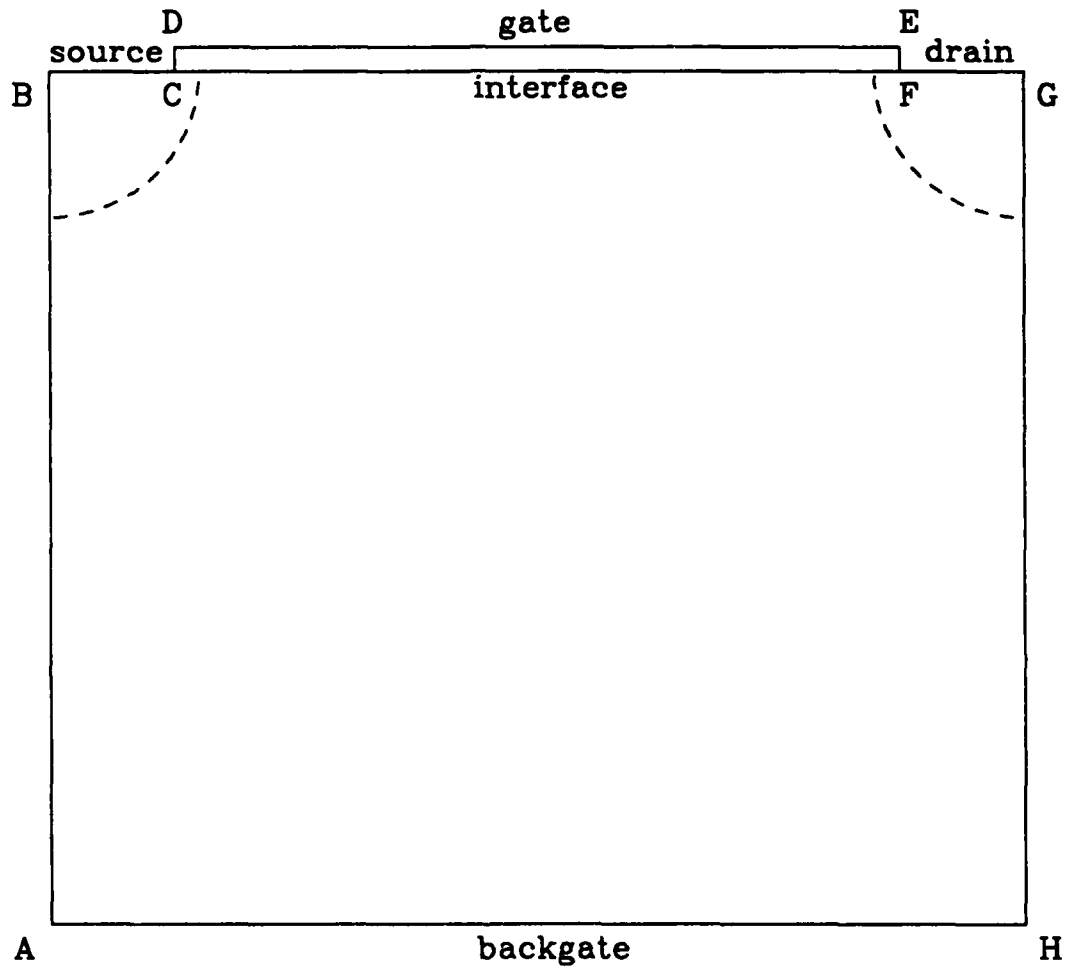
k_1 is the net doping density.

The current for electrons \mathbf{J}_n in the semiconductor region

$$\mathbf{J}_n = q\mu_n n \mathbf{E} + qD_n \nabla n,$$

is the sum of the drift current $q\mu_n n \mathbf{E}$ and the diffusion current $qD_n \nabla n$, where

MOSFET geometry



\mathbf{E} is the electric field $-\nabla\psi$,
 μ_n is the electron mobility,
 D_n is the electron diffusion coefficient.

For holes the current

$$\mathbf{J}_p = q\mu_p p \mathbf{E} - qD_p \nabla p,$$

is the sum of the drift current $q\mu_p p \mathbf{E}$ and the diffusion current $-qD_p \nabla p$, where

μ_p is the hole mobility,
 D_p is the hole diffusion coefficient.

The continuity equations for the electrons and the holes provide the second and third partial differential equations

$$-\nabla \cdot \mathbf{J}_n - q(G - R) + q \frac{\partial n}{\partial t} = 0,$$

and

$$\nabla \cdot \mathbf{J}_p - q(G - R) + q \frac{\partial p}{\partial t} = 0,$$

where

G introduces generation phenomena,
 R introduces recombination processes.

We can remove a few constants from our equations by rephrasing them as follows: Instead of the electrostatic potential ψ we use the dimensionless potential $u = \frac{q\psi}{k_B T}$, where

k_B is Boltzmann's constant,
 T is temperature.

The densities n , p and k_1 are expressed in units of the equilibrium concentration or intrinsic density n_i of the semiconductor [11]. We assume that the dielectric constants ϵ_{xx} and ϵ_{zz} are isotropic and uniform. Spatial dimensions are measured in terms of the intrinsic Debye length

$$L_i = \sqrt{\frac{\epsilon_s \epsilon_0 k_B T}{q^2 n_i}}.$$

Within the semiconductor region Ω the equations for the steady state problem can thus be rewritten in terms of "natural," physical coordinates and by introducing the function k_2 for describing recombination and generation phenomena as

$$\begin{aligned} f_1(u, n, p) &= -\nabla^2 u + n - p - k_1 = 0, \\ f_2(u, n, p) &= -\nabla \cdot \left(-\frac{k_B T}{q} \mu_n n \nabla u + D_n \nabla n \right) + k_2 = 0, \\ f_3(u, n, p) &= \nabla \cdot \left(-\frac{k_B T}{q} \mu_p p \nabla u - D_p \nabla p \right) + k_2 = 0. \end{aligned}$$

Boundary conditions. The potential u is defined on the entire region Ω' , which includes the thin oxide layer on top. The densities n and p are defined only on the main region Ω which consists of the major quadrangle.

For the potential u on Ω' the boundary conditions are given by:

- At source and drain the Dirichlet condition

$$u = u_{\text{contact}} + \log\left\{\frac{k_1}{2} + \sqrt{\left(\frac{k_1}{2}\right)^2 + 1}\right\},$$

where u_{contact} is the applied electrostatic potential u_{source} or u_{drain} and the additional term is generated by the doping k_1 .

- At the gate the Dirichlet condition that u is equal to the applied constant u_{gate} . Because there is no static space-charge in the oxide between the gate and the semiconductor, the potential within the oxide satisfies Laplace's equation:

$$-\nabla \cdot (\epsilon_{ox} \epsilon_0 \nabla u) = 0,$$

where ϵ_{ox} is the relative dielectric constant of the oxide. To the sides of the oxide we have the homogeneous Neumann condition $\nabla u \cdot \mathbf{n} = 0$, corresponding to no perpendicular electric field. Across the oxide-semiconductor interface we require continuity of the potential u and continuity of the normal electric field:

$$\epsilon_s \frac{\partial u_s}{\partial \mathbf{n}} = \epsilon_{ox} \frac{\partial u_{ox}}{\partial \mathbf{n}}.$$

- Along the sides the homogeneous Neumann condition $\nabla u \cdot \mathbf{n} = 0$ because the electric field perpendicular to the side boundaries is zero.
- At the backgate at the back of the device the constant Dirichlet condition that $u = 0$. Thus the backgate is our point of reference for the potential u .

For the densities n and p on Ω the boundary conditions are given by:

- At source, drain and backgate the Dirichlet condition

$$n = \frac{k_1}{2} + \sqrt{\left(\frac{k_1}{2}\right)^2 + 1},$$

with the corresponding condition for p , which follows from the equality $np = 1$, which is known to be valid under thermal equilibrium from solid state physics [18]. The expression for n follows from $n - p - k_1 = 0$, expressing space-charge neutrality ($\rho = 0$) at the drain and source contacts.

- At the oxide-semiconductor interface the Neumann conditions

$$\mathbf{J}_n \cdot \mathbf{n} = (-k_B T \mu_n n \nabla u + q D_n \nabla n) \cdot \mathbf{n} = 0,$$

$$\mathbf{J}_p \cdot \mathbf{n} = (-k_B T \mu_p p \nabla u - q D_p \nabla p) \cdot \mathbf{n} = 0,$$

where \mathbf{n} is unit vector normal to the interface. This means that no current can flow across the interface.

- At the sides the Neumann conditions

$$\mathbf{J}_n \cdot \mathbf{n} = (-k_B T \mu_n n \nabla u + q D_n \nabla n) \cdot \mathbf{n} = 0,$$

$$\mathbf{J}_p \cdot \mathbf{n} = (-k_B T \mu_p p \nabla u - q D_p \nabla p) \cdot \mathbf{n} = 0,$$

because no current is supposed to flow across these boundaries either.

3. Analysis of the system of partial differential equations

In this section we show that it follows readily from the semiconductor equations that the physical charge is conserved while certain maximum principles hold. Charge conservation is expressed mathematically (Jackson [7]) in terms of the physical current density \mathbf{J} ($\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p$) as this density having zero divergence:

$$\nabla \cdot \mathbf{J} = 0.$$

A second important property follows if Einstein's relations (Kittel [11])

$$D_n = \mu_n \frac{k_B T}{q},$$

$$D_p = \mu_p \frac{k_B T}{q},$$

relating the diffusivities D pointwise to the mobilities μ , are assumed to hold. Making this assumption we obtain certain maximum principles for the solution in terms of the coordinates u, ν and ω , where the coordinates ν and ω are defined in terms of n and p in terms of the natural coordinates of the preceding section, by

$$\nu = e^{-u} n,$$

$$\omega = e^u p.$$

The situation is as summarized in the following

Theorem 3.1. For the solution u, n, p to the system of mixed boundary value problems

$$\begin{aligned} f_1(u, n, p) &= -\nabla^2 u + n - p - k_1 = 0, \\ f_2(u, n, p) &= -\nabla \cdot \mathbf{J}_n + qk_2 = 0, \\ f_3(u, n, p) &= \nabla \cdot \mathbf{J}_p + qk_2 = 0, \end{aligned} \quad (3.1)$$

where

$$\begin{aligned} \mathbf{J}_n &= -k_B T \mu_n n \nabla u + q D_n \nabla n, \\ \mathbf{J}_p &= -k_B T \mu_p p \nabla u - q D_p \nabla p, \end{aligned}$$

subject to the boundary conditions as described in the preceding section,

- Charge is conserved in the sense mentioned above.
- Assuming Einstein's relations as mentioned above and introducing the coordinates ν and ω , the electron and hole currents can be rewritten as

$$\begin{aligned} \mathbf{J}_n &= -\mu_n k_B T n \nabla u + q D_n \nabla n = \mu_n k_B T e^u \nabla \nu, \\ \mathbf{J}_p &= -\mu_p k_B T p \nabla u - q D_p \nabla p = -\mu_p k_B T e^{-u} \nabla \omega. \end{aligned} \quad (3.2)$$

Hence with respect to these coordinates we obtain the equations

$$\begin{aligned} h_1(u, \nu, \omega) &= -\nabla^2 u + e^u \nu - e^{-u} \omega - k_1 = 0, \\ h_2(u, \nu, \omega) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^u \nabla \nu \right) + k_2 = 0, \\ h_3(u, \nu, \omega) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_p e^{-u} \nabla \omega \right) + k_2 = 0, \end{aligned} \quad (3.3)$$

and if $k_2 = 0$ the solutions ν and ω have to assume their maximum at the boundary $\delta\Omega$.

In the theorem above the potential equation $\nabla \cdot (\epsilon_{ox} \nabla u) = 0$ for the oxide region is implicitly included in the first of the equations (3.1), (3.3).

Proof.

From the second and third partial differential equation in (3.1) we observe readily that the divergence of the conduction electron and hole current densities \mathbf{J}_n and \mathbf{J}_p are opposite: The total current $\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p$ satisfies

$$\nabla \cdot \mathbf{J} = \nabla \cdot (\mathbf{J}_n + \mathbf{J}_p) = 0.$$

Thus we see that charge is conserved.

For the proof of the second part of the theorem we assume that Einstein's relations are valid. With respect to the coordinates ν and ω which we introduced in this section, the current continuity equations then exhibit particularly strong regularity properties.

First of all, the identities for the conduction electron and hole currents (3.2) mentioned in the theorem follow straightforwardly.

Moreover, if $k_2 = 0$, we now have obtained current continuity equations for which the conditions for Theorem 8.1 in [5] are satisfied, and therefore the solutions ν and ω have to assume their maximum at the boundary $\delta\Omega$.

Finally it may be observed that because each of the partial differential equations which constitute the system is in divergence form, there exists a weak formulation of the problem (confer [5], chapter 8). Similar systems have been analyzed by Seidman in [17]. Jerome has shown in [8] that the system describing carrier transport in semiconductors possesses a weak solution with gradients bounded in the L_2 norm, where the L_2 norm of the gradient is a special case of the more general L_q norm, which for $q > 0$ in two or three dimensions is defined by:

$$\|\nabla u\|_{L_q} = \left[\int |\nabla u|^q dx \right]^{\frac{1}{q}} \text{ where } n = 2, 3.$$

4. Analysis of the discretized equations

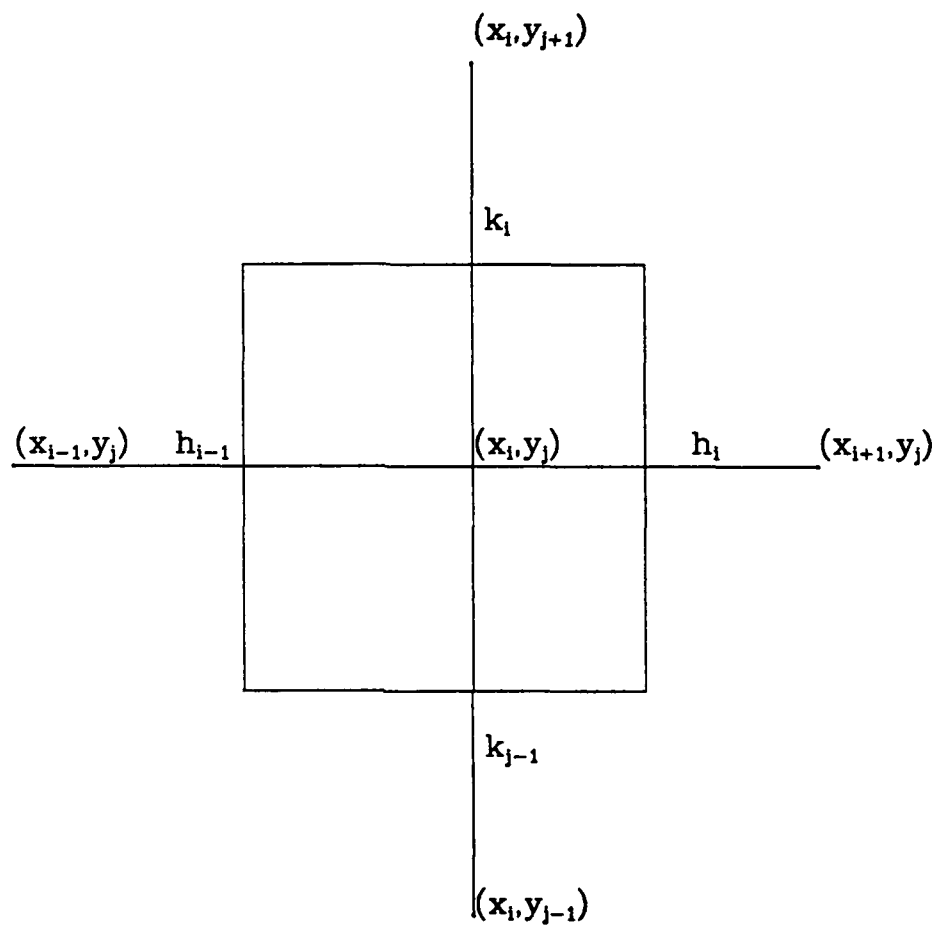
The systems of equations which the numerical solution vectors have to satisfy are usually obtained from the system of pdes by finite elements or finite differences. Because the oxide is extremely thin for typical MOSFET devices, we do not generate any extra gridpoints or unknowns within the oxide. For the points at the interface we obtain the correct equations using interface conditions.

As conservation of charge and the maximum principles for ν and ω are important properties of the solution, we discretize in such a way that discrete counterparts hold for the discrete solution.

We define conservation of charge for finite difference equations by introducing a discretized version of the equation $\nabla \cdot \mathbf{J} = 0$. More specifically, we obtain the charge-conserving finite difference equations by exploiting the divergence form of the current continuity equations $\nabla \cdot \mathbf{J} = qk$. To this end we employ the box-scheme as described in Varga [19] or in [1], and Gauß's law (see figure 2) for forming the discrete equations. In order to obtain the discrete equation at an interior point (x_i, y_j) corresponding to $\nabla \cdot \mathbf{J} = qk$ we first write

$$\int \int_{box} \nabla \cdot \mathbf{J} dx dy = \int \int_{box} qk dx dy \Rightarrow$$

Finite difference scheme



$$\int_{\partial \text{box}} \mathbf{J} \cdot d\sigma = \int_{\text{top}} \mathbf{J} \cdot d\sigma + \int_{\text{right}} \mathbf{J} \cdot d\sigma + \int_{\text{bottom}} \mathbf{J} \cdot d\sigma + \int_{\text{left}} \mathbf{J} \cdot d\sigma = \int \int_{\text{box}} qk dx dy,$$

Hence we see from the divergence form of the differential operator and by writing $I_s = \int_s \mathbf{J} \cdot d\sigma$ (where s is top, right, bottom or left) for the current flux through any of the four sides, that for the continuous problem

$$I_{\text{top}} + I_{\text{right}} + I_{\text{bottom}} + I_{\text{left}} = \int \int_{\text{box}} qk dx dy, \quad (4.1)$$

where the right hand side is 0 if the divergence of the current is 0.

Now we can discretize by approximating each of the fluxes I_{top} , I_{right} , I_{bottom} and I_{left} through the sides of the box by a finite difference approximation \tilde{I}_{top} , \tilde{I}_{right} , $\tilde{I}_{\text{bottom}}$ and \tilde{I}_{left} , and approximating the right hand side $\int \int_{\text{box}} qk dx dy$ by $q\tilde{k}$.

We then say that charge is conserved for the discrete equations if $\tilde{k} = 0$ in

$$\tilde{I}_{\text{top}} + \tilde{I}_{\text{right}} + \tilde{I}_{\text{bottom}} + \tilde{I}_{\text{left}} = q\tilde{k}.$$

Hence for discretized equations, charge conservation can be obtained from charge conservation for the original pde in divergence form if the pde is discretized by employing expression (4.1) and subsequently approximating the four current fluxes using finite differences.

In Theorem 3.1 we obtained, under the assumption of Einstein's relations, for the unknowns u , ν and ω the equations

$$h_1(u, \nu, \omega) = -\nabla^2 u + e^u \nu - e^{-u} \omega - k_1 = 0,$$

$$h_2(u, \nu, \omega) = -\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^u \nabla \nu \right) + k_2 = 0,$$

$$h_3(u, \nu, \omega) = -\nabla \cdot \left(\frac{k_B T}{q} \mu_p e^{-u} \nabla \omega \right) + k_3 = 0.$$

We saw moreover that the components ν and ω of the solution to this system of pdes, satisfy certain maximum principles. For the discrete solution vectors corresponding to ν and ω the discrete maximum principle means that the extremal values of the components of the solution vectors have to be assumed at those components which correspond to boundary points.

We consider the case for gridpoint (x_i, y_j) (see figure 2) where

$$h_i = x_{i+1} - x_i,$$

$$k_j = y_{j+1} - y_j$$

and

$$\tilde{h}_i = \frac{1}{2}(h_{i-1} + h_i),$$

$$\tilde{k}_j = \frac{1}{2}(k_{j-1} + k_j).$$

The value of a component of a solution vector v for point (x_i, y_j) we write down as $v_{i,j}$.

As a counterpart to Theorem 3.1 we now summarize how the properties of the discrete system can reproduce those of the original pdes in the following

Theorem 4.1. Consider the discrete current continuity equations at point (x_i, y_j) , and let $\tilde{I}_{n,top}$, $\tilde{I}_{n,right}$, $\tilde{I}_{n,bottom}$ and $\tilde{I}_{n,left}$ be any finite difference expressions for the discretized fluxes through the sides of the discretization box above, while $\tilde{I}_{p,top}$, $\tilde{I}_{p,right}$, $\tilde{I}_{p,bottom}$ and $\tilde{I}_{p,left}$ are the corresponding fluxes for holes. Then

- If identical discretization procedures are used for conduction electrons and holes, charge is conserved with respect to these finite difference approximations to the fluxes in the sense that

$$\tilde{I}_{n,top} + \tilde{I}_{n,right} + \tilde{I}_{n,bottom} + \tilde{I}_{n,left} + \tilde{I}_{p,top} + \tilde{I}_{p,right} + \tilde{I}_{p,bottom} + \tilde{I}_{p,left} = 0.$$

- If we assume that Einstein's relations are valid we can discretize the fluxes through the sides of the box such that the matrices A for the current continuity equations satisfy that
 - All off-diagonal elements on each row of A have the same sign.
 - The diagonal elements of A each have the opposite sign the off-diagonal elements have on that row.
 - $|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|$.

and therefore we obtain discrete maximum principles in the sense mentioned above for the solution vectors corresponding to the unknowns v and w , if the generation-recombination terms are 0.

Proof. The discretized current continuity equations are obtained by employing expression (4.1) and subsequently approximating the four current fluxes I_n for conduction-electrons and I_p for holes using finite differences. The current continuity equation for conduction electrons thus yields the equation

$$\tilde{I}_{n,top} + \tilde{I}_{n,right} + \tilde{I}_{n,bottom} + \tilde{I}_{n,left} = q\tilde{k}_2,$$

while the current continuity equation for holes results in

$$\tilde{I}_{p,top} + \tilde{I}_{p,right} + \tilde{I}_{p,bottom} + \tilde{I}_{p,left} = -q\tilde{k}_2.$$

By adding these two equations for the finite difference approximations to the fluxes we see that at any interior point of the grid

$$\tilde{I}_{n,top} + \tilde{I}_{n,right} + \tilde{I}_{n,bottom} + \tilde{I}_{n,left} + \tilde{I}_{p,top} + \tilde{I}_{p,right} + \tilde{I}_{p,bottom} + \tilde{I}_{p,left} = 0.$$

Thus discretized charge is conserved at (x_i, y_j) if we discretize both current continuity equations according to (4.1) and use identical finite difference approximations for both conduction-electron and hole currents.

At the points adjacent to a boundary where a Dirichlet condition is imposed the same analysis is valid while for points on a boundary where we have a Neumann condition some fluxes can be 0 but everything else is unchanged.

Employing the geometrical definitions introduced before the theorem, the statement concerning the maximum principle is an adaptation from Theorem 6.1.1 in Ortega [14]. This theorem states that the solution to a linear system $Ax = b$ satisfies a discrete maximum principle if (Ortega [14])

- All off-diagonal elements on each row of A have the same sign.
- The diagonal elements of A each have the opposite sign the off-diagonal elements have on that row.
- $|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|$,

and we have a right hand side which is appropriately 0. We can satisfy these conditions if we approximate the first derivatives for the current fluxes in the previously mentioned box scheme in such a way that each of them generates a positive contribution to the diagonal coefficient which is the opposite of the corresponding off-diagonal coefficient. Because the current continuity equations are pdes which are of the form $\nabla \cdot (a \nabla \nu)$ while the function a satisfies $a > 0$, this can be done.

■

The last point in the proof is illustrated by the following example: For the equations for ν we could discretize the first derivatives with respect to ν such that we obtain

$$\begin{aligned} & \tilde{I}_{n,i,j+\frac{1}{2}} + \tilde{I}_{n,i+\frac{1}{2},j} + \tilde{I}_{n,i,j-\frac{1}{2}} + \tilde{I}_{n,i-\frac{1}{2},j} = \\ & -\frac{\tilde{h}_i}{k_j} \mu_n k_B T \frac{1}{2} (e^{u_{i,j+1}} + e^{u_{i,j}}) (\nu_{i,j+1} - \nu_{i,j}) - \frac{\tilde{k}_j}{h_i} \mu_n k_B T \frac{1}{2} (e^{u_{i+1,j}} + e^{u_{i,j}}) (\nu_{i+1,j} - \nu_{i,j}) \\ & -\frac{\tilde{h}_i}{k_{j-1}} \mu_n k_B T \frac{1}{2} (e^{u_{i,j-1}} + e^{u_{i,j}}) (\nu_{i,j-1} - \nu_{i,j}) - \frac{\tilde{k}_j}{h_{i-1}} \mu_n k_B T \frac{1}{2} (e^{u_{i-1,j}} + e^{u_{i,j}}) (\nu_{i-1,j} - \nu_{i,j}) = \\ & -\tilde{h}_i \tilde{k}_j q k_{2,i,j}. \end{aligned}$$

If $k_{2,i,j} = 0$ we can conclude from the cited theorem that a discrete maximum principle is valid for the discrete solution ν . An analogous situation holds for ω .

5. Choice of coordinates for the discretized equations

In the previous sections we have seen that the desire to conserve charge and to preserve certain maximum principles for the solution to the discrete equations determine a number of aspects of the discretization process. Thus far, we have not yet considered the properties of a given set of equations with regard to numerical solution. In fact, further transformations are necessary so as to obtain well-conditioned numerical problems.

First of all, we still have to linearize the nonlinear systems in order to solve them. Subsequently the linear algebraic systems of equations have to be solved repeatedly so as to solve the nonlinear difference equations.

The choice of both the system of coordinates for discretization and for linearization therefore determines important aspects of the linear systems which have to be solved. We now analyze the merits of a number of coordinate systems for numerical solution of the equations, following approximately the approach of Bank, Rose and Fichtner in [1].

One option for solving the systems of nonlinear equations is employing Newton's method for determining the search direction at each successive iteration, and do an inexact line-search along this direction. Thus quadratic convergence will be retrieved once the iterates are sufficiently close to the solution that no damping is needed. This is attractive in the sense that only a limited number of nonlinear iterations will be necessary.

Although this is a fast nonlinear algorithm, it requires the solution of the full Newton equations

$$Jdx = -f(x),$$

at each iteration, which is relatively expensive.

We form the solution vector for the entire system by taking the part of the vector which corresponds to the potential u first, and placing the two components of the solution which correspond to density-like parameters in the second and third part of the vector. Likewise we take the discretized equations for the potential equation first and the equations corresponding to the current continuity equations as second and third set. With respect to this ordering the block zero-structure of this Jacobian J follows from the functional dependencies of the three pdes: The first equation for the potential involves both the densities n and p as well as the potential u , and each of the current continuity equations depends on u as well as its own specific unknown.

Hence the Jacobian for Newton's method is a 3 by 3 block matrix, provided we do not reorder equations and unknowns between blocks after discretizing and linearizing. From the relations between the pdes mentioned above it follows that the symbolical block-zero-structure of the matrix

is given by

$$J = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & \\ A_{31} & & A_{33} \end{pmatrix}.$$

Preferably Newton's equations should be solved by a fast, iterative algorithm. Only iterative methods can take full advantage of the extreme sparsity of the matrices resulting from 3-dimensional problems, the ultimate object in semiconductor simulation.

It does not even appear to be feasible, however, to discretize and linearize in such a way that we will obtain a full Jacobian which has properties like symmetric positive definiteness, which are most conducive to the rapid convergence of fast iterative methods, such as conjugate gradients.

Because of the difficulty of the solution of the full Newton equations, a number of decoupling algorithms have been devised in which only systems of equations which correspond roughly to the diagonal blocks have to be solved.

A large class of nonlinear decoupling algorithms is collectively denoted by "Gummel's method" and goes back to an iterative scheme which was proposed by H.K. Gummel in 1964 [6] for one dimensional simulation. In this approach at each iteration, each of the three systems of, possibly nonlinear, equations is solved successively for the corresponding variables while the other two sets are kept fixed. The iteration terminates once the three components of the solution are sufficiently consistent.

The typical theoretical decoupling results [13] or, under physically more realistic regularity conditions, [9] and [10] all predict convergence for algorithms in which the current continuity equations are solved for density-like unknowns like ν and ω in section 3 and 4 or a function thereof. It turns out that even in those cases the algorithms will converge only for limited variation of the boundary data, however. For the Jacobian this corresponds to domination of the A_{21} and A_{31} blocks by corresponding diagonal blocks.

As we shall see, it is possible to find discretization and linearization coordinates such that these diagonal A_{11} , A_{22} and A_{33} blocks of the Jacobian, which have to be solved only in a decoupled approach, are symmetric positive definite, and thus amenable to solution by a rapidly converging iterative algorithm. However, the range of the size of the components of the solution vectors corresponding to the current continuity equations for these symmetric representations is so large that it does not allow sufficiently accurate solution in finite precision arithmetic except for small variation of the boundary conditions.

Alternatively, it is sometimes attempted to solve the full Newton equations by a block Gauß-Seidel or block SOR approach [3] [4], in which only the diagonal blocks have to be solved. Thus a block-iterative scheme is used for the solution of the linear systems of equations for Newton's method. However, just like the off-diagonal coupling has to be small for convergence of nonlinear decoupling algorithms like Gummel's, the off-diagonal coupling has to be small as well for the convergence of these block iterative methods for the solution of the linear 3 by 3 block Jacobian, such as block-Jacobi, block Gauß Seidel or block-SOR. Thus there seems to be no reason to expect the block iterative methods to converge for the linearized problem if the nonlinear decoupling algorithm will not converge or vice versa. At best we may expect the three diagonal blocks to be valuable as a preconditioning for an iterative method for the solution of the full Jacobian.

Such preconditionings will be more effective if the diagonal blocks of the Jacobian dominate the off-diagonal blocks sufficiently. The extent to which this is the case depends on the choice of coordinates as well. Thus it is important to formulate the equations in an expedient coordinate system.

We will proceed to examine how the properties of the systems of coupled nonlinear equations and of the linearized systems depend on the choice of variables. Because of considerations concerning conditioning, charge conservation, maximum principles and meshwidth restrictions we will recommend only two alternative sets of coordinates. Other formulations will be shown to be either almost completely useless or useful only for restricted classes of problems.

6. Adjusted densities as coordinates

In Theorem 3.1 we have demonstrated that under the assumption that Einstein's relations are valid, the equations describing transport in semiconductor devices with respect to the coordinates u, ν and ω are given by

$$\begin{aligned} h_1(u, \nu, \omega) &= -\nabla^2 u + e^u \nu - e^{-u} \omega - k_1 = 0, \\ h_2(u, \nu, \omega) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^u \nabla \nu \right) + k_2 = 0, \\ h_3(u, \nu, \omega) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_p e^{-u} \nabla \omega \right) + k_2 = 0. \end{aligned}$$

Thus the current continuity equations are linear with respect to their own variable and the solutions ν and ω satisfy maximum principles.

We represent the Jacobian for this system symbolically without indicating the discretization procedure as in [1]. The symbol $*$ denotes a so-called "placeholder" for the solution vector. Hence $\nabla^2 *$ corresponds to some matrix which coincides with a discretized Laplacean and scalar terms

indicate diagonal terms of a matrix. Thus the Jacobian for this system is equal to

$$J = \begin{pmatrix} -\nabla^2 * + e^u \nu + e^{-u} \omega & e^u & -e^{-u} \\ -\nabla \cdot (\frac{k_B T}{q} \mu_n e^u * \nabla \nu) & -\nabla \cdot (\frac{k_B T}{q} \mu_n e^u \nabla *) & \\ \nabla \cdot (\frac{k_B T}{q} \mu_p e^{-u} * \nabla \omega) & & -\nabla \cdot (\frac{k_B T}{q} \mu_p e^{-u} \nabla *) \end{pmatrix}.$$

The difference equations are derived by the procedure advocated in Section 4. Thus we obtain first equations like (4.1) for the fluxes through the sides of the discretization box again. Subsequently these fluxes are approximated by finite difference expressions. We can approximate the conduction-electron fluxes through the sides of the boxes employed in the discretization scheme for instance by

$$I_{n,i,j+\frac{1}{2}} = -\frac{\tilde{h}_i}{k_j} \mu_n k_B T \frac{1}{2} (e^{u_{i,j+1}} + e^{u_{i,j}}) (\nu_{i,j+1} - \nu_{i,j}),$$

or (Scharfetter-Gummel)

$$I_{n,i,j+\frac{1}{2}} = -\frac{\tilde{h}_i}{k_j} \mu_n k_B T \frac{\frac{1}{2}(u_{i,j+1} - u_{i,j})}{\sinh(\frac{1}{2}(u_{i,j+1} - u_{i,j}))} e^{u_{i,j+1}+u_{i,j}} (\nu_{i,j+1} - \nu_{i,j}),$$

etc. Motivations for discretizing as proposed by Scharfetter and Gummel are physical rather than mathematical and are given in [3], for instance. As stated in Theorem 4.1, by employing either of these discretizations charge is conserved while the desired maximum principles are valid.

Nevertheless, the current formulation of the problem is not very well suited for numerical purposes. For instance, for a typical simulation problem the boundary values for the potential vary over 5 Volt. Because the electric potentials are multiplied by the factor $\frac{q}{k_B T} \approx 38.68$, however, 5 Volt corresponds to a dimensionless potential u of 198.4. This results in $e^u \approx 2.69 \cdot 10^{43}$. Therefore, the occurrence of the coefficients e^u makes the full Jacobian or even just its diagonal blocks insufferably ill conditioned and even too large for representation in most floating point number systems. Moreover, the variables ν and ω can be shown to become relatively large in parts of the domain Ω . This makes the off-diagonal A_{21} and A_{31} blocks large with respect to the corresponding diagonal blocks which does not coincide with good conditioning. Therefore this is not a set of coordinates which is very well suited for actual numerical computations.

Thus, we have to make further coordinate transformations because of ill-conditioning of the discrete equations. In particular we should try to choose our ultimate set of coordinates such that only differences of the potential u occur as arguments for the function $x \rightarrow e^x$ in our calculations.

The variable ν is a coordinate which is too evasive because its range is too large. Whereas the physical density $n = e^{u-\nu}$ might vary over a range of 10^{30} , it can still be represented in most

floating point number representations. The range of $\nu = ne^{-u}$, however, can cause severe accuracy difficulties because it may cover a range of 10^{80} due to the multiplication by e^{-u} .

Note that this is not just a problem because of the range of floating point number systems, but much more so because of the relative machine precision. Solving linear systems for a solution of which the smallest components are evanescent with respect to the largest ones leaves no accuracy for these smallest components. Obtaining the density n from ν by multiplying by e^u , these large relative errors in small components of the solution vector produce large relative inaccuracies in large components.

The ill-conditioning of the diagonal blocks for the second and third equation is so severe that we hardly ever want to use this representation for calculations although these blocks are symmetric positive definite and the equations are linear with respect to their own variable.

7. Natural coordinates

As observed in the preceding section, we can obtain systems of equations with symmetric coefficient matrices for the vectors for u, ν and ω , but these are generally not suited for numerical solution. From these illconditioned systems we can obtain equations for u, n and p which are better conditioned, however, by scaling on the right hand side. This amounts to first discretizing the current continuity equations with respect to the variables ν and ω as an intermediate step and subsequently incorporating factors e^{-u} and e^u as a scaling on the right. Thus the maximum principles for ν and ω are preserved. In this approach we solve the equations

$$\begin{aligned} h_1(u, n, p) &= -\nabla^2 u + n - p - k_1 = 0, \\ h_2(u, n, p) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^u \nabla e^{-u} n \right) + k_2 = 0, \\ h_3(u, n, p) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_p e^{-u} \nabla e^u p \right) + k_2 = 0, \end{aligned}$$

in u, n and p , which can be discretized so as to preserve charge, to have solutions which adhere to the desired maximum principles and such that the numerical properties are more agreeable.

It is not so important that the A_{22} and A_{33} blocks which we obtain this way are nonsingular. The theoretical analysis of the system of pdes does not yield any decoupling results with respect to u, n and p anyway.

Proceeding as above we obtain the Jacobian

$$J = \begin{pmatrix} -\nabla^2 u & 1 & -1 \\ -\nabla \cdot \left[\frac{k_B T}{q} \mu_n e^u (*\nabla e^{-u} n - \nabla e^{-u} n*) \right] & -\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^u \nabla e^{-u} * \right) & \\ +\nabla \cdot \left[\frac{k_B T}{q} \mu_p e^{-u} (*\nabla e^u p - \nabla e^u p*) \right] & & -\nabla \cdot \left(\frac{k_B T}{q} \mu_p e^{-u} \nabla e^u * \right) \end{pmatrix}.$$

Because the coordinates for which we solve are the actual physical variables u , n and p , the numerical solution of this formulation of the problem will be feasible as long as these physical variables can be represented in the number system which is used. Although this set of coordinates is probably not well suited for decoupled approaches, it has been employed successfully for a full Newton approach, e.g. [2].

The last two diagonal blocks are no longer positive definite in the sense that $\langle v, Av \rangle \geq 0$ need no longer hold, but they do have a positive spectrum. This follows because the matrix corresponding to

$$\nabla \cdot (e^u \nabla *) e^{-u}$$

is similar to the symmetric positive definite matrix corresponding to

$$e^{-\frac{u}{2}} \nabla \cdot (e^u \nabla *) e^{-\frac{u}{2}}$$

as can be seen from

$$\nabla \cdot (e^u \nabla *) e^{-u} = e^{\frac{u}{2}} [e^{-\frac{u}{2}} \nabla \cdot (e^u \nabla *) e^{-\frac{u}{2}}] e^{-\frac{u}{2}}.$$

Here $e^{-\frac{u}{2}} \nabla \cdot (e^u \nabla *) e^{-\frac{u}{2}}$ is symmetric positive definite because it is symmetric and congruent to the positive definite matrix corresponding to $\nabla \cdot (e^u \nabla *)$.

8. Quasi Fermi levels as coordinates

Expressing densities in units of the intrinsic carrier density n_i of the semiconductor, we can transform from the variables u , n and p to the coordinates u , v and w defined by

$$n = e^{u-v},$$

$$p = e^{w-u}.$$

These v and w are called quasi-Fermi levels and as they are logarithmic coordinates for the densities they may be expected to be smoother and to have a smaller range.

Assuming that Einstein's relations

$$D_n = \mu_n \frac{k_B T}{q},$$

$$D_p = \mu_p \frac{k_B T}{q},$$

are valid again, the electron and hole currents can be rewritten as

$$\mathbf{J}_n = -\mu_n k_B T n \nabla u + q D_n \nabla n = \mu_n k_B T e^{u-v} \nabla v$$

and

$$\mathbf{J}_p = -\mu_p k_B T p \nabla u - q D_p \nabla p = -\mu_p k_B T e^{w-u} \nabla w.$$

Expressing the currents thus in terms of the coordinates u, v and w the current continuity equations can be simplified so as to obtain

$$\begin{aligned} g_1(u, v, w) &= -\nabla^2 u + e^{u-v} - e^{w-u} - k_1 = 0, \\ g_2(u, v, w) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^{u-v} \nabla v \right) + k_2 = 0, \\ g_3(u, v, w) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_p e^{w-u} \nabla w \right) + k_2 = 0. \end{aligned}$$

As before, μ_n and μ_p are arbitrary positive mobility functions. Because the current continuity equations are again in divergence form we can again discretize in such a way that for the discrete equations charge is conserved. For v and w we can obtain discrete maximum principles because we have equations of the form $\nabla \cdot (a \nabla v)$ which have to be solved for v while the function a satisfies $a > 0$.

The current continuity equations obtained this way are each nonlinear with respect to their own variable. However, they still possess a unique solution. For the second equation we have, for instance

$$\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^{u-v} \nabla v \right) = -k_2 \Leftrightarrow \nabla \cdot \left(\frac{k_B T}{q} \mu_n e^u \nabla e^{-v} \right) = -k_2.$$

Hence any unique solution v to the equation

$$\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^u \nabla v \right) = -k_2,$$

corresponds to exactly one solution v of the first equation and vice versa, with the relation between them given by $v = e^{-v}$.

For discretization we can employ the box scheme again. After linearizing in these same variables u, v and w , the Jacobian needed for employing Newton's method is given by

$$J = \begin{pmatrix} -\nabla^2 * + e^{u-v} + e^{w-u} & -e^{u-v} & -e^{w-u} \\ -\nabla \cdot \frac{k_B T}{q} \mu_n e^{u-v} * \nabla v & -\nabla \cdot \frac{k_B T}{q} \mu_n e^{u-v} (- * \nabla v + \nabla *) & \\ \nabla \cdot \frac{k_B T}{q} \mu_p e^{w-u} * \nabla w & & -\nabla \cdot \frac{k_B T}{q} \mu_p e^{w-u} (* \nabla w + \nabla *) \end{pmatrix}.$$

The differential operators in this Jacobian are not that amenable for obtaining linear systems with favourable properties because v and w occur both as arguments of exponents and as linear factors. This complicates managing the linear algebraic properties of many of the blocks. Therefore we want to obtain the linear systems from a different system of nonlinear equations.

More agreeable nonlinear equations in terms of quasi Fermi levels can be obtained from the equations (3.3) by writing $\nu = e^{-v}$ and $\omega = e^w$. The logarithms of ν and ω are well-defined because we have shown in Theorem 4.1 on the properties of discrete solutions, that the solution vectors ν and ω to the current continuity equations are unique and satisfy discrete maximum principles. Hence the entire solution vectors ν and ω are larger than 0 because the boundary conditions for both ν and ω are larger than 0.

We thus obtain nonlinear equations for v and w from the linear equations for ν and ω . From the maximum principle for the discretized $\nu = e^{-v}$ and $\omega = e^w$, maximum principles for the discrete v and w follow trivially because of the monotonicity of $x \rightarrow e^x$ and $x \rightarrow e^{-x}$.

For the conduction electrons we thus employ a discretization according to $h_2(u, e^{-v}, e^w) = -\nabla \cdot (\frac{k_B T}{q} \mu_n e^u \nabla e^{-v}) + k_2 = 0$ as shown in section 4 and we can obtain, for instance, the equations

$$\begin{aligned} & \frac{\tilde{h}_i}{k_j} \mu_n e^{\frac{u_{i,j+1} + u_{i,j}}{2}} (e^{-v_{i,j+1}} - e^{-v_{i,j}}) + \frac{\tilde{k}_j}{h_i} \mu_n e^{\frac{u_{i+1,j} + u_{i,j}}{2}} (e^{-v_{i+1,j}} - e^{-v_{i,j}}) \\ & + \frac{\tilde{h}_i}{k_{j-1}} \mu_n e^{\frac{u_{i,j-1} + u_{i,j}}{2}} (e^{-v_{i,j-1}} - e^{-v_{i,j}}) + \frac{\tilde{k}_j}{h_{i-1}} \mu_n e^{\frac{u_{i-1,j} + u_{i,j}}{2}} (e^{-v_{i-1,j}} - e^{-v_{i,j}}) = \tilde{h}_i \tilde{k}_j \frac{q}{k_B T} k_{2,i,j}. \end{aligned}$$

In order to improve the conditioning of this system of equations which is linear with respect to ν , we multiply by $e^{-u_{i,j}}$ on the left, and form a system of equations which is nonlinear in ν by dividing the i -th row of the system $h_2(u, \nu, \omega) = 0$ by ν_i .

That this nonlinear system of equations in ν is equivalent to the linear one we started out with, we prove in the next

Lemma 8.1. *Let A be a nonsingular N by N matrix. Then every vector ν which satisfies $\nu_i \neq 0 \forall i = 1, \dots, N$, and which solves either the linear system*

$$\sum_{j=1}^N A_{ij} \nu_j = r_i, \quad i = 1, \dots, N, \quad (8.1)$$

or the nonlinear system

$$\sum_{j=1}^N A_{ij} \frac{\nu_j}{\nu_i} = \frac{r_i}{\nu_i}, \quad i = 1, \dots, N, \quad (8.2)$$

solves the other system of equations as well and is unique.

Proof. Let ν be a vector for which $\nu_i \neq 0 \forall i = 1, \dots, N$. If ν satisfies (8.1) then ν is the unique solution to a linear system of equations for which the coefficient matrix is nonsingular. Moreover, because $\nu_i \neq 0 \forall i = 1, \dots, N$, (8.2) follows from (8.1). Hence ν solves the system of nonlinear equations.

If, the other way around, ν satisfies the nonlinear equations (8.2), it has to satisfy (8.1) because $\nu_i \neq 0 \quad \forall \quad i = 1, \dots, N$. The solution to the nonlinear equations is unique because the solution to the linear equations is unique.

The coefficient matrices for the linear systems of equations for ν and ω are nonsingular and the discrete vectors ν and ω are larger than 0. Therefore Lemma 8.1 is applicable and we have a unique solution for the appropriate nonlinear equations in the sense of this lemma.

Multiplying by e^{-u} and dividing through by the solution vector ν is probably only advisable under the assumption that $k_2 = 0$ because otherwise the original right hand side of the pdes becomes dependent on u and v or w .

By all the above transformations we thus obtain in terms of u and v the equations

$$\begin{aligned} & \frac{\tilde{h}_i}{k_j} \mu_n e^{\frac{u_{i,j+1}-u_{i,j}}{2}} (e^{v_{i,j}-v_{i,j+1}} - 1) + \frac{\tilde{k}_j}{h_i} \mu_n e^{\frac{u_{i+1,j}-u_{i,j}}{2}} (e^{v_{i,j}-v_{i+1,j}} - 1) \\ & + \frac{\tilde{h}_i}{k_{j-1}} \mu_n e^{\frac{u_{i,j-1}-u_{i,j}}{2}} (e^{v_{i,j}-v_{i,j-1}} - 1) + \frac{\tilde{k}_j}{h_{i-1}} \mu_n e^{\frac{u_{i-1,j}-u_{i,j}}{2}} (e^{v_{i,j}-v_{i-1,j}} - 1) = 0. \end{aligned}$$

By Lemma 8.1 we still have a unique solution with nonzero components e^{-v} to this nonlinear system, which is identical to the solution of the linear system of equations in terms of ν . From the uniqueness of the solution for $\nu = e^{-v}$ we trivially obtain uniqueness for the solution v .

The Jacobian $J_{22} = (j_{ij})$ for the linearized systems with respect to v gives rise to equations in the Newton update δv with rows

$$\begin{aligned} & -\frac{\tilde{h}_i}{k_j} \mu_n e^{\frac{u_{i,j+1}-u_{i,j}}{2}} e^{v_{i,j}-v_{i,j+1}} \delta v_{i,j+1} - \frac{\tilde{k}_j}{h_i} \mu_n e^{\frac{u_{i+1,j}-u_{i,j}}{2}} e^{v_{i,j}-v_{i+1,j}} \delta v_{i+1,j} \\ & - \frac{\tilde{h}_i}{k_{j-1}} \mu_n e^{\frac{u_{i,j-1}-u_{i,j}}{2}} e^{v_{i,j}-v_{i,j-1}} \delta v_{i,j-1} - \frac{\tilde{k}_j}{h_{i-1}} \mu_n e^{\frac{u_{i-1,j}-u_{i,j}}{2}} e^{v_{i,j}-v_{i-1,j}} \delta v_{i-1,j} \\ & + \left[\frac{\tilde{h}_i}{k_j} e^{\frac{u_{i,j+1}-u_{i,j}}{2}} e^{v_{i,j}-v_{i,j+1}} + \frac{\tilde{k}_j}{h_i} e^{\frac{u_{i+1,j}-u_{i,j}}{2}} e^{v_{i,j}-v_{i+1,j}} \right. \\ & \left. + \frac{\tilde{h}_i}{k_{j-1}} e^{\frac{u_{i,j-1}-u_{i,j}}{2}} e^{v_{i,j}-v_{i,j-1}} + \frac{\tilde{k}_j}{h_{i-1}} e^{\frac{u_{i-1,j}-u_{i,j}}{2}} e^{v_{i,j}-v_{i-1,j}} \right] \delta v_{i,j}. \end{aligned}$$

Thus the Jacobian for the second equation is always irreducibly row-diagonally dominant with a positive diagonal and negative off-diagonal elements. Therefore it cannot become singular so that Newton's method may be expected to converge rapidly.

In fact, the Jacobian J_{22} with respect to v is a M-matrix because

- (i) for the diagonal elements we have $j_{ii} > 0$,

- (ii) the off-diagonal elements satisfy $j_{kl} < 0 (k \neq l)$,
- (iii) the matrix J_{22} is irreducibly diagonally dominant,

([14], Theorem 6.2.17.) Therefore the incomplete LU factorization of J is well defined ([12], Theorem 2.3.) Thus we have an effective preconditioning at our disposal when solving this block with an iterative method.

Both for solution of the full Jacobian and for a decoupling algorithm, quasi Fermi levels are thus attractive coordinates because they have a small range, we can discretize in such a way that we obtain the desired discrete maximum principles, charge can be conserved for the discrete equations and because we can rewrite the equations in such a way that we only have differences of potentials, differences of potentials and quasi Fermi levels and differences of quasi Fermi levels as arguments of the exponential function. Both the group at Bell labs [1] and the group at Philips [15] appear to have employed these coordinates, although neither group has given specific details on the implementation.

9. Hybrid approaches

We have by no means exhausted all the systems of coordinates in which the problem can be represented, or all the possible discretizations. It should be observed, that apart from employing different coordinates for discretization of the differential operators and for linearization, it is possible to change coordinates after having linearized.

Changing coordinates after having linearized so as to remove scaling problems of the Jacobian will usually not be very effective, however. The ill-scaled solution vector corresponding to the Jacobian itself will have to be recovered from a solution to a transformed problem for the inexact line-searches along the Newton directions which are necessary. Therefore it only makes sense to transform coordinates in order to improve the conditioning of the linear systems if we do not generate the entire Jacobian, but use some nonlinearly decoupled algorithm like Gummel's method, for instance.

The decoupling algorithms are defined in terms of the potential u and the quasi Fermi levels v and w or their exponents ν and ω . Nevertheless we can employ, for instance, the coordinates n and p in an intermediate step for obtaining the next iterates from the current continuity equations. An algorithm in terms of u, v and w with intermediate coordinates n and p is thus given by

$$\begin{aligned} g_1(u^l, v^l, w^l) &= -\nabla^2 u^l + e^{u^l-v^l} - e^{w^l-u^l} - k_1 = 0, \\ h_2(u^l, n^l, p^l) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_n e^{u^l} \nabla e^{-u^l} n^{l+1} \right) + k_2 = 0, \end{aligned}$$

$$\begin{aligned}
h_3(u^l, n^l, p^l) &= -\nabla \cdot \left(\frac{k_B T}{q} \mu_p e^{-u^l} \nabla e^{u^l} p^{l+1} \right) + k_2 = 0, \\
v^{l+1} &= u^l - \log(n^{l+1}), \\
w^{l+1} &= u^l + \log(p^{l+1}), \\
l &= l + 1.
\end{aligned}$$

Hence the symmetry of the current continuity equations in terms of ν and ω is sacrificed in favor of the conditioning. This way we only have to solve linear systems for the current continuity equations.

For the decoupling algorithms the problem of ill-conditioning for the symmetric positive definite current continuity equations can be ameliorated somewhat while still preserving symmetric positive definiteness, by forming the equations which correspond to $e^{-\frac{u}{2}} \nabla \cdot (e^u \nabla e^{-\frac{u}{2}} *)$. Thus we rewrite the A_{22} and A_{33} blocks for the (u, ν, ω) coordinate-system, in terms of the coordinates u , $\tilde{\nu} = e^{-\frac{u}{2}} n$ and $\tilde{\omega} = e^{\frac{u}{2}} p$. Subsequently we can multiply A_{22} on the left by $e^{-\frac{u}{2}}$ and A_{33} by $e^{\frac{u}{2}}$. This decoupling algorithm is given by

$$\begin{aligned}
-\nabla^2 u^l + e^{u^l} \nu^l - e^{-u^l} \omega^l - k_1 &= 0, \\
-e^{-\frac{u^l}{2}} \nabla \cdot (\mu_n e^{u^l} \nabla e^{-\frac{u^l}{2}} \tilde{\nu}^{l+1}) &= 0, \\
-e^{\frac{u^l}{2}} \nabla \cdot (\mu_p e^{-u^l} \nabla e^{\frac{u^l}{2}} \tilde{\omega}^{l+1}) &= 0, \\
\tilde{\nu}^{l+1} &= e^{\frac{u^l}{2}} \nu^{l+1}, \\
\tilde{\omega}^{l+1} &= e^{-\frac{u^l}{2}} \omega^{l+1}, \\
l &= l + 1.
\end{aligned}$$

Hence symmetric systems are obtained for the current continuity equations which can be solved efficiently using conjugate gradients. Thus whenever the boundary conditions are so mild that $\tilde{\nu}$ and $\tilde{\omega}$ can be used as coordinates, it is to be recommended for a decoupled approach.

It may appear that we can obtain more favourable equations with respect to the quasi Fermi levels, for instance, if we differentiate symbolically before we discretize. Thus we can eliminate the exponentials from the last two equations under the assumption that $\mu_n = \mu_p = 1$ and $k_2 = 0$. The nonlinear equations are then given by

$$\begin{aligned}
g_1(u, v, w) &= -\nabla^2 u + e^{u-v} - e^{w-u} - k_1 = 0, \\
\hat{g}_2(u, v, w) &= -\nabla^2 v + |\nabla v|^2 - \nabla u \cdot \nabla v = 0, \\
\hat{g}_3(u, v, w) &= -\nabla^2 w - |\nabla w|^2 + \nabla u \cdot \nabla w = 0.
\end{aligned}$$

Linearizing in u , v and w we obtain the Jacobian

$$J = \begin{pmatrix} -\nabla^2 * + e^{u-v} + e^{w-u} & -e^{u-v} & -e^{w-u} \\ -\nabla v \cdot \nabla * & -\nabla^2 * + \nabla(2v - u) \cdot \nabla * & \\ \nabla w \cdot \nabla * & & -\nabla^2 w + \nabla(u - 2w) \cdot \nabla * \end{pmatrix}.$$

Thus the last two diagonal blocks are neither positive definite, nor are they symmetric. For the second and third equation the off-diagonal blocks are smaller than the diagonal blocks, however. Although the diagonal blocks continue to be nonsymmetric, the second and third equation are now only quadratic and no longer exponential. The solution to these nonlinear equations is again unique because of a similar argument as we gave for the case of quasi Fermi levels.

Major objections to this kind of approach are, however, that neither charge is conserved, nor discrete maximum principles are valid in general. Moreover, singularity problems can accrue from the first derivative terms in a decoupled approach if the mesh is not extremely fine. Hence eliminating exponentials from equations by formally differentiating and dividing them out does not appear to be a fruitful approach.

Many alternative sets of coordinates and ways of discretizing which are thus still conceivable are therefore not mentioned here because it is not advisable to use them.

10. Conclusion

The solution to the original system of pdes has two important properties (Theorem 3.1):

1. Physical charge is conserved.
2. In the absence of recombination and generation the coordinates ν and ω have to assume their extremal values at the boundary.

Hence we want to discretize in such a way that the discrete counterparts to these characteristics are valid for the discrete equations.

Two major approaches are available for solution of the system of nonlinear equations

- For solution of the system by a decoupling algorithm
 1. u, ν, ω result in linear symmetric positive definite systems for the current continuity equations, but these are illconditioned for all but near trivial problems (Section 6). Hence just the linear systems can be scaled further, but this improves conditioning only marginally if symmetry is to be preserved. Giving up symmetry and solving for u, n and p while thereafter transforming to u, v, w results in well-conditioned linear current continuity equations (Section 7).

2. u, n, p are not suitable. Even for trivial boundary data there are no theoretical results indicating that decoupling occurs for these coordinates.
 3. u, v, w is a suitable set of coordinates which results in three systems of nonlinear equations (Section 8).
- For solution of the system by a full Newton approach
 1. u, ν, ω do not result in a well conditioned Jacobian. This set of coordinates gives rise to symmetric positive definite systems for the diagonal blocks for the current continuity equations, but this does not simplify solution of the full Jacobian (Section 6).
 2. u, n, p result in a Jacobian which is not very well conditioned, but which appears to be manageable still (Section 7).
 3. u, v, w is a suitable set of coordinates which results in a Jacobian for which the off diagonal blocks can be limited (Section 8).

References

- [1] Randolph E. Bank, Donald J. Rose, Wolfgang Fichtner, *Numerical methods for semiconductor device simulation*, SIAM Journal on Scientific and Statistical Computing, 4/3 September (1983), pp. 416-435.
- [2] E.M. Buturla, P.E. Cottrell, *Simulation of Semiconductor Transport Using Coupled and Decoupled Solution Techniques*, Solid State Electronics, 23 July (1980), pp. 331-334.
- [3] Wolfgang Fichtner, Donald J. Rose, Randolph E. Bank, *Semiconductor device simulation*, SIAM Journal on Scientific and Statistical Computing, 4/3 September (1983), pp. 391-415.
- [4] Andrea F. Franz, Gerhardt A. Franz, Siegfried Selberherr, Christian Ringhofer, Peter Markowich, *Finite Boxes-A Generalization of the Finite-Difference Method Suitable of Semiconductor Device Simulation*, IEEE transactions on electron devices, ED-30/9 September (1983), pp. 1070-1082.
- [5] David Gilbarg and Neil S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag, 1983.
- [6] H.K. Gummel, *A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor calculations*, IEEE transactions on electron devices, ED-11 (1964), pp. 455-465.
- [7] J.D. Jackson, *Classical Electrodynamics (2nd edition)*, John Wiley and Sons, 1974.
- [8] Joseph W. Jerome, *Consistency of Semiconductor Modelling: An Existence/Stability Analysis for the Stationary van Roosbroeck System*, SIAM J. Appl. Math., 45/4 August (1985).
- [9] ———, *The Role of Semiconductor Device Diameter and Energy-Band Bending in Convergence of Picard Iteration for Gummel's Map*, IEEE Trans. Elec. Dev., (To be published).
- [10] Tom Kerkhoven, *On the dependence of the convergence of Gummel's algorithm on the regularity of the solution*, To be published.
- [11] Charles Kittel, *Introduction to solid state Physics (5th edition)*, John Wiley and Sons, 1976.
- [12] J.A. Meijerink and H. A. van der Vorst, *An Iterative Solution Method for Linear Systems of Which the Coefficient Matrix is a Symmetric M-Matrix*, Mathematics of Computation, 31/137(1977), pp. 148-162.
- [13] M.S. Mock, *On Equations Describing Steady-State Carrier Distributions in a Semiconductor Device*, Comm. Pure Appl. Math., 25 (1972), pp. 781-792.
- [14] James M. Ortega, *Numerical Analysis, A Second Course*, Academic Press, 1972.
- [15] S.J. Polak, A. Wachters, H.M.J. Vaes, A de Beer and C. den Heijer, *An Algorithm for the Calculation of Poisson's Equation in 2-D Semiconductor problems*, Int.J.Num.Meth.Eng., 17 (1981), pp. 1591-1604.

- [16] W. van Roosbroeck, *Theory of the flow of electrons and holes in Germanium and other semiconductors*, Bell Sys. Tech. J., /29 (1950), pp. 560.
- [17] Thomas I. Seidman, *Steady State Solutions of Diffusion-Reaction Systems with Electrostatic Convection*, Nonlinear Analysis. Theory, Methods and Applications, 4 (1980), pp. 623-637.
- [18] S.M. Sze, *Physics of Semiconductor Devices (2nd Edition)*, Wiley-Interscience, 1981.
- [19] R.S. Varga, *Matriz Iterative Analysis*, Prentice-Hall, Englewood Cliffs, 1962.

END

FILMED

5-85

DTIC